



Balancing California's Grid Without Batteries

Neil Cammardella, Joel Mathias, Matthew Kiener, Ana Bušić, Sean Meyn

► To cite this version:

Neil Cammardella, Joel Mathias, Matthew Kiener, Ana Bušić, Sean Meyn. Balancing California's Grid Without Batteries. 57th IEEE Conference on Decision and Control (CDC 2018), Dec 2018, Miami, United States. hal-01968606

HAL Id: hal-01968606

<https://hal.science/hal-01968606>

Submitted on 2 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Balancing California's Grid Without Batteries

Neil Cammardella¹, Joel Mathias¹, Matthew Kiener², Ana Bušić³, and Sean Meyn¹

Abstract—*Demand dispatch* refers to a collection of distributed control techniques to obtain grid services from flexible loads. A carefully designed control architecture can enable a collection of loads to behave in aggregate as a large virtual storage device. Grid-level ancillary services can be provided with minimal communication, while guaranteeing quality of service to the consumer.

This work expands on prior work in several directions:

- A natural notion of *energy capacity* is proposed for the special case of thermostatically controlled loads (TCLs). It is shown that this quantity is closely approximated by thermal energy capacity, which is a component of the “leaky battery model” introduced in prior work.
- Simulation experiments in a distributed control setting show that these energy limits, and accompanying power capacity limits, are reliable indicators of online capacity, even for a heterogeneous population of loads.
- A feedforward/feedback control scheme is proposed for a large collection of heterogeneous loads. At the local level, control loops are used to create cooperative responses from each load in a given class of homogeneous loads. This simplifies control of the aggregate based on two pieces of information: aggregate power consumption from each class of loads and the *state of charge* surrogate that is a part of the leaky battery model. This information is required at a slow time-scale (say, 5 minute sampling).
- The paper concludes with economic implications. In particular, given that dispatchable loads are a form of virtual storage and not virtual generation, it is not surprising that the use of real time or time-of-use pricing has been problematic.

I. INTRODUCTION

California witnessed several records in March of 2018. The California Independent System Operator (CAISO) saw an all-time peak percentage of demand served by solar, hitting a record of nearly 50% at 1pm. The next day, CAISO set another solar record, this time hitting a new peak for solar production of over 10 GW in the morning. Prices are negative during these events, and the system suffers from physical shocks in terms of massive ramping in supply, as well as a large peak in energy consumption in the afternoon. The magnitude of these shocks is evident in Fig. 1, which shows the large penetration of renewables, especially solar,

Funding from the Department of Energy Building Technologies Office (GMLC grant *virtual batteries*), and our partners at PNNL and ORNL are gratefully acknowledged. Additional funding was provided through the ARPA-E NODES award, *Open and Scalable Distributed Energy Resource Networks*, National Science Foundation award EPCN 1609131, and French National Research Agency grant ANR-16-CE05-0008.

¹Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611

²Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, 94305

³Inria and DI ENS, École Normale Supérieure, CNRS, PSL Research University, Paris, France

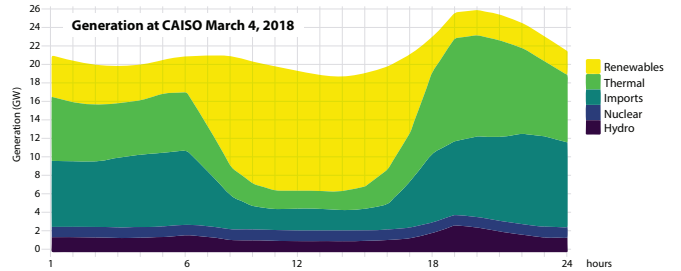


Fig. 1. Record solar in California — March 4, 2018.

in the CAISO region. The afternoon ramp in demand from thermal generation is about 5 GW per hour; for reference, a large nuclear power plant is about 1.5 GW.

Government policy has been enacted in various regions of the globe to encourage the installation of resources to smooth energy from renewables. In California there are incentives / mandates for installation of massive battery systems. The largest lithium-ion battery system in the world was installed there in 2017, only to be out-done by a slightly larger installation in Australia a few months later [1]. Among the challenges with batteries:

- Energy loss of about 10% per charge/discharge cycle.
- Air-conditioning may be required to cool the fleet.
- It is estimated that the battery cells must be replaced every 3-5 years, depending on usage.
- Large amount of real-estate is required for current systems (see images in [1]).
- The installation cost is enormous [2].

In parallel with these policy decisions is a rapidly evolving science for *demand dispatch*: utilization of the inherent flexibility of many electric loads to create the same services that can be offered by batteries. This *virtual energy storage* (VES) is not intended to “store energy for a rainy day”. Rather, just like the battery systems that are installed today, flexible loads will virtually “charge and discharge” in response to a command signal from a grid operator or balancing authority (BA). The goals of the BA will be met: ramps from solar generation will be tempered, and afternoon peak demand will be leveled. In our design, the quality of service (QoS) offered to the consumer is kept within strict bounds.

Not all battery systems are alike. For example, lithium-ion batteries are slow to charge, but can discharge quickly. Nickel metal-hydrate technology has a more symmetric charge/discharge rate. The same is true of virtual energy storage: the characteristics of VES from a collection of residential water heaters is different from a collection of

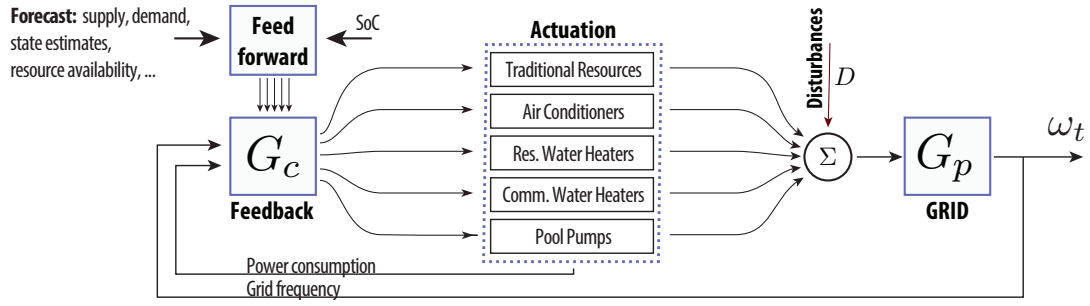


Fig. 2. The control architecture: feedforward commands are computed from forecasts and feedback commands are computed from real-time error.

commercial units. Consequently, the integration of heterogeneous resources is an important component in the dispatch of storage – either real or virtual.

The solution to this problem is naturally addressed through a feedforward/feedback architecture. The feedforward part is essential to take advantage of forecasts of grid events, such as ramps in solar or wind energy production, and the approximately periodic demand cycle. In particular, in anticipation of a peak in baseline net-load, the population of water heaters should heat to near the endpoint of the deadband interval determined by the thermostat. This is analogous to charging a battery.

Feedback is required because forecasts are not perfect, especially for high frequency volatility induced by renewables or traditional generation outages.

The implementation of feedforward control is based on Model Predictive Control (MPC), which requires a dynamical model as well as constraints. The dynamics of virtual energy storage through the distributed control architecture of [3], [4] gives rise to an input-output model that is typically passive, but complex. This complexity has been addressed in later work: in [5], [6] it is argued that an additional local control loop can be introduced at each load to make the aggregate input-output dynamics approximately linear, and also flat within a specified bandwidth. By adopting this distributed control strategy, the only model required is a surrogate “state of charge” (SoC) for each class of loads. A useful model of the SoC for VES from thermostatically controlled loads (TCLs) was proposed in [7], [8] (a component of the “leaky battery model”), and for residential pools in [3], [4].

A block diagram of the proposed control strategy is shown in Fig. 2, whose components are described broadly as follows:

- (i) The block G_p (GRID) captures dynamics such as the grid inertia and frequency-dependent loads in the network controlled by the BA. Linear models can be found in standard references [9].
- (ii) The “actuators” represent the collection of bundled loads; each class providing VES with certain characteristics.
- (iii) The feedforward block determines optimal trajectories for each of the resources — VES as well as batteries and traditional generation. This optimization step is repeated periodically (every 30 minutes in the numerical

examples considered). This computation requires estimates of the state of charge at the start of the optimization computation, and forecasts of load and renewable generation.

(iv) The feedback block serves two roles:

- *Slow time scale.* Based on the desired signals obtained from the feedforward block, and measurements of aggregate power, a scalar command signal is synthesized and broadcast to each load class to ensure that the signals are tracked.
- *Fast time scale.* This is similar to automatic generation control (AGC). Regulation of the grid frequency ω_t is obtained using a feedback architecture that is typical in practice today. A PI control architecture is suitable, given the simple grid dynamics, and the flat dynamics of the batteries or VES resources. The scalar command signals may be decomposed by frequency, as described in [10], [6].

In this paper it is assumed that the BA has access to two types of information from each class of loads: aggregate SoC and aggregate power consumption. It is argued in [5], [6] that control on the fast time scale can be achieved using only grid-level frequency measurements.

The present paper is most closely related to [11], which considers the feedforward control problem under the assumption that the histograms of loads is available at the BA.

The SoC constraints of [7], [8] are essentially constraints on thermal energy storage, which is not of direct interest to the grid operator. In Prop. 2.1 it is shown that the thermal energy bound obtained in this prior work is closely approximated by electrical energy capacity. See [12] for discussion on the multi-dimensional aspects of energy capacity for VES.

Net-present-value comparisons of real and virtual energy storage can be found in [13], [14].

The remainder of this paper is organized as follows. Section II estimates the power and energy capacities of TCLs and validates those estimates. Section III describes the distributed control architecture. Simulations of flexible loads providing grid services are performed in IV. Conclusions and directions for future research are discussed in V.

II. VIRTUAL STORAGE: POWER AND ENERGY CAPACITY

In this section, we describe how virtual energy storage can be provided by thermostatically controlled loads (TCLs). We restrict our attention to water heaters, though the models

are developed similarly for other thermal loads such as refrigerators and air conditioners¹.

Consider a collection of N water heaters indexed by $k = \{1, 2, \dots, N\}$. The *state process* for the k -th load at time t is defined as the pair

$$X^k(t) = (M^k(t), \Theta^k(t)), \quad (1)$$

in which $M^k(t) \in \{0, 1\}$ denotes the power mode (the value “1” indicating the unit is on), and $\Theta^k(t)$ the inside temperature of the tank; all temperatures are in units of degree Fahrenheit.

A. Individual water heater model

Focusing on an individual load, we drop the superscript in (1). The basic function of the water heater is to keep the internal water temperature $\Theta(t)$ at time t within the (time-invariant) lower and upper bounds as $\Theta_- \leq \Theta(t) \leq \Theta_+$. The standard model relating these state variables and disturbances is a first order ODE:

$$\frac{d}{dt}\Theta(t) = -\lambda(\Theta(t) - \Theta_a) + \gamma M(t)\mathcal{P}_m - (\Theta(t) - \Theta_{in}(t))\nu f(t) \quad (2)$$

where Θ_a is the (time-invariant) ambient air temperature, $\Theta_{in}(t)$ is temperature of the cold water entering the tank, $f(t)$ is flow rate of hot water from the unit (gallons/sec), $M(t) \in \{0, 1\}$ is the power mode ($M(t) = 1$ indicates that the power is on), \mathcal{P}_m is the on-power, and $\lambda, \gamma, \nu > 0$ are scalar parameters governing heat transfer of the system.

Simple formulae for power and energy capacity are obtained in an idealized setting, as in [7], [8], in which there is no water is moving in and out of the tank:

$$\frac{d}{dt}\Theta(t) = -\lambda(\Theta(t) - \Theta_a) + \gamma M(t)\mathcal{P}_m. \quad (3)$$

It is assumed that the parameters are such that the water will cool down ($\frac{d}{dt}\Theta(t) < 0$) when the power is off ($M(t) = 0$), and heat up ($\frac{d}{dt}\Theta(t) > 0$) when the power is on ($M(t) = 1$). Sufficient conditions for these are $\Theta_- > \Theta_a$ and $\gamma\mathcal{P}_m > \lambda(\Theta_+ - \Theta_a)$, respectively.

The dynamics of a basic hysteresis controller are shown in Fig. 3. When the water temperature reaches the upper limit, the heater turns off. The water then becomes cooler until it reaches the lower limit, at which point the heater turns on. The heater off-duration T_{off} and on-duration T_{on} are computed from equation (3) as

$$\lambda T_{off} = \ln\left(1 + \frac{\Theta_+ - \Theta_-}{\Theta_- - \Theta_a}\right), \quad (4a)$$

$$\lambda T_{on} = \ln\left(1 + \frac{\Theta_+ - \Theta_-}{\gamma\mathcal{P}_m/\lambda - (\Theta_+ - \Theta_a)}\right). \quad (4b)$$

The average power over time is given by as $\mathcal{P}_0 = \mathcal{P}_m \frac{T_{on}}{T_{on} + T_{off}}$.

¹For cooling devices, the sign conventions will be different in the models.

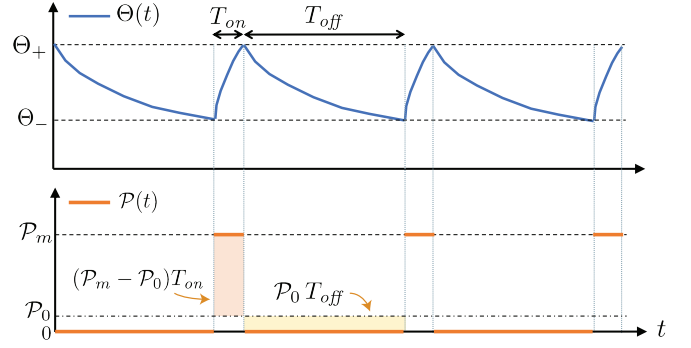


Fig. 3. Dynamics of water temperature and power under a water tank hysteresis controller.

B. Power capacity

The notion of capacity in terms of power for a collection of homogeneous loads is defined in [8]. the upper and lower power limits are denoted η_+ and η_- . In a stationary setting in which the average power of the collection is meaningful, it follows that $\eta_+ = N\mathcal{P}_0$ and $\eta_- = N(\mathcal{P}_m - \mathcal{P}_0)$.

The quantity $N\mathcal{P}_0$ is regarded as the baseline power consumption. If we turn off all the loads, then this is analogous to a battery discharging at level η_+ . Turning on all loads is similar to a battery charging at rate η_- .

This definition of power capacity can be easily extended to the case with hot water usage by replacing the average power consumption by a baseline power consumption for the aggregate.

C. Thermal energy capacity

The energy bound obtained in [8] is based on their formulation of *state of charge* (SoC) of a homogeneous collection of TCLs. This is the aggregate temperature of a collection of loads, normalized so that it is expressed in units of energy, with a constant subtracted so the nominal value is zero. Consequently, the SoC evolves as a first order linear system similar to (3):

$$\frac{d}{dt}x(t) = -\lambda x(t) - u(t), \quad (5)$$

in which $u(t)$ is the aggregate power deviation (relative to a baseline) supplied to the grid by the collection of loads. It is interpreted as the power drawn from the VES composed by the collection of loads. The SoC is subject to the bound $|x(t)| \leq E^0/2$ for all t , where E^0 is the thermal energy limit. It follows from Corollary 7 of [8] that the thermal energy limit for a homogeneous collection of N WHs is

$$E^0 = N(\Theta_+ - \Theta_-)/\gamma \quad (6)$$

D. Energy capacity

The energy capacity of a homogeneous collection of loads is defined in terms analogous to that of an energy storage system such as a battery; i.e., the energy capacity is the amount of energy that can be delivered by the storage system during a discharge period.

For a homogeneous collection of N loads, the discharge period is equal to the discharge period of an individual load.

A “fully charged” collection of N water heaters corresponds to each individual temperature at the maximal value Θ_+ . The maximal “discharge period” of the population coincides with T_{off} of an individual (see Fig. 3).

Virtual energy storage capacity for the collection of N homogeneous loads is average power times a discharge period: $E = N\mathcal{P}_0 \times T_{off}$.

It is demonstrated next that this capacity limit is approximately consistent with the energy capacity bound (6).

The error term (2.1) appearing in Prop. 7 is typically very small, since in practice $\Theta_- - \Theta_a \gg \Theta_+ - \Theta_-$. For example, in the case of water heaters, $\Theta_- - \Theta_a$ is typically between 40–60 °F, whereas the temperature deadband $\Theta_+ - \Theta_-$ varies from 3–7 °F.

Proposition 2.1: The following approximation holds for a homogeneous collection of N water heaters without usage:

$$E = E^0 + O(N\mu^2), \quad \mu = \frac{\Theta_+ - \Theta_-}{\Theta_- - \Theta_a} \quad (7)$$

Proof: Denote the average temperature by

$$\Theta_0 := \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \Theta(t) dt.$$

The average power consumption for one load can be expressed in the two equivalent forms:

$$\mathcal{P}_0 = \mathcal{P}_m \frac{T_{on}}{T_{off} + T_{on}} = \frac{\lambda}{\gamma} (\Theta_0 - \Theta_a).$$

where the second equation follows from (3). The expression $E = N\mathcal{P}_0 \times T_{off}$ then yields

$$E = N \frac{\lambda}{\gamma} (\Theta_0 - \Theta_a) \times T_{off} \quad (8)$$

A Taylor series approximation gives

$$\begin{aligned} T_{off} &= \frac{1}{\lambda} \frac{\Theta_+ - \Theta_-}{\Theta_- - \Theta_a} + O(\mu^2) \\ &= \frac{1}{\lambda} \frac{\Theta_+ - \Theta_-}{\Theta_0 - \Theta_a} + O(\mu^2) \end{aligned}$$

Substituting this bound into (8) gives (7). ■

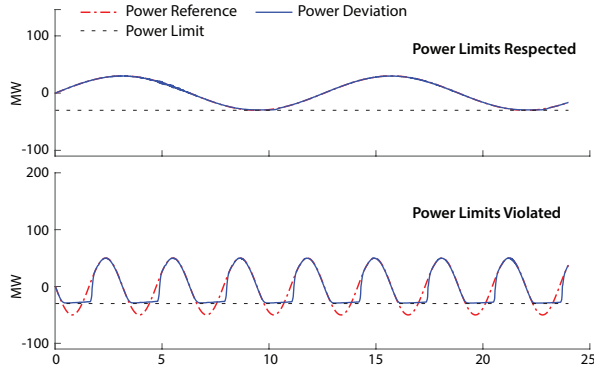


Fig. 4. *Respecting power limits:* The upper figure shows excellent tracking. The lower figure shows poor tracking when the discharge power limit is violated. (Charge power limit is out of scale.)

E. Realizing capacity

A homogeneous collection of 120,000 water heaters was simulated under the myopic design of [6]. This distributed control architecture is based on local randomized control at each load. A common discrete-time command signal is sent to all water heaters, then each water heater considers both the command signal and its own temperature to compute the probability of changing the power mode.

Simulations were performed with a variety of reference signals. Figs. 4 and 5 show aggregate power deviation from nominal for 120,000 water heaters. Note that the power signals shown in Figs. 4–7 are the power consumed by the collection of loads relative to a baseline (i.e., $-u(t)$ in (5)).

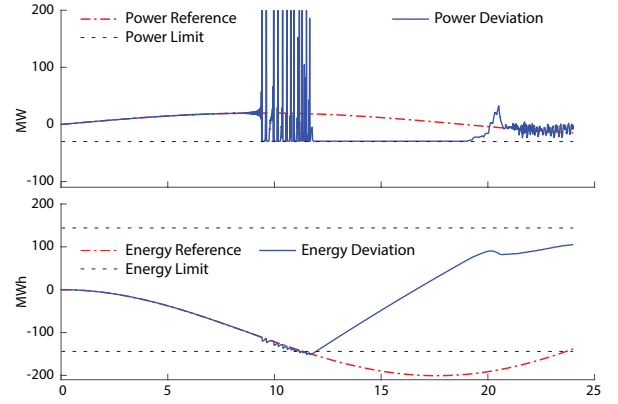


Fig. 5. *Respecting energy limits:* Poor tracking occurs when the energy capacity is exceeded.

Fig. 4 shows excellent tracking of a 30 MW, 120 MWh signal. However, when the discharge power exceeds 30 MW, poor tracking occurs. In general, tracking is excellent when the reference signal respects the power and energy constraints of the loads.

Fig. 5 demonstrates poor tracking when the reference signal exceeds the energy limit of the collection of water heaters. Observe that instability occurs when the SoC is within about 90% of the energy limit. Maintaining such high ranges of the SoC require a very high rate of cycling from the individual loads; cycling limits enforced by the distributed control algorithm result in the oscillations observed in the figure.

The energy capacity limits were tested in more realistic settings that included non-stationary usage (details of the simulation model can be found in [6]). The three rows in Fig. 6 correspond to three scalings of a grid reference signal. In the first subplot, the reference signal is zero (the plot in the upper left hand corner shows nominal aggregate power consumption in this case). In the final row, the reference signal violates power limits, resulting in poor tracking between hours 12–17.

Also shown in the figure is the behavior of one typical water heater among the fleet in each of the three cases. The qualitative behavior of power consumption is virtually unchanged, even when the reference signal exceeds power limits.

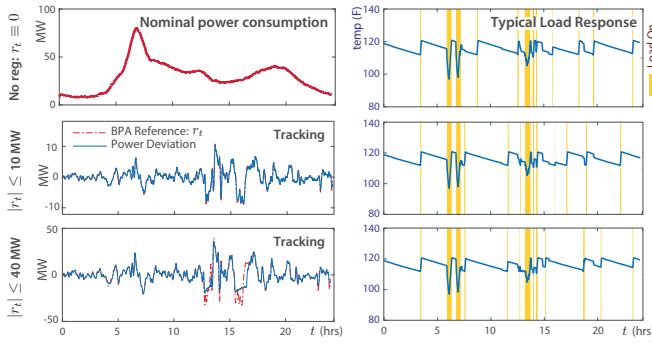


Fig. 6. Power capacity experiments with heterogeneous water heaters with usage. The morning peak in nominal power consumption is consistent with typical water usage.

A second set of experiments were performed to investigate the impact of energy limits. Results are shown in Fig. 7. An unstable response is observed when the SoC approaches the energy capacity $E^0/2$. The tracking is excellent when energy limits are not violated.

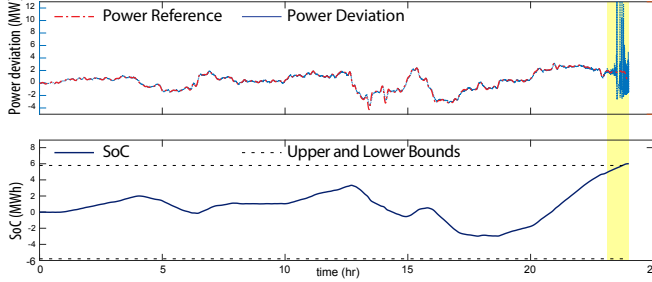


Fig. 7. Energy capacity with heterogeneous water heaters with usage.

III. CONTROL ARCHITECTURE

Experience indicates that it is often advantageous to combine feedback and feedforward, and the correct balance requires insight and understanding of their respective properties [15]. In our own prior work we have ignored the feedforward component, although we have pointed out the obvious drawbacks [5]. In particular, reliance on pure real-time control excludes the possibility of “pre-charging” virtual energy storage resources in anticipation of peaks in net load. The focus of this work is on the feedforward architecture, which operates on a slow time-scale (e.g., updates every 30 minutes). The feedback component has been covered extensively in previous work [5], [6], [16].

The control models are described here in continuous time.

A. TCL model

The generalized battery model (5) is used for each class of loads. The SoC x and power deviation u are subject to the following dynamic constraints, for each $t \geq 0$:

$$\frac{d}{dt}x(t) = -\alpha x(t) - u(t) \quad (9)$$

$$-\eta_- \leq u(t) \leq \eta_+ \quad (10)$$

$$|x(t)| \leq C, \quad (11)$$

in which $C = NE^0/2$, and η_{\pm} are defined in Section II-D.

B. Residential swimming pools

Similar abstract notions of SoC were introduced in [3], [4] for a collection of residential pool pumps. For an individual load with state process \mathbf{X} evolving in discrete time, it is assumed that there is a “snapshot” QoS function $\ell(\mathbf{X}(k))$, and that total QoS at time τ is the discounted sum:

$$\mathcal{L}(\tau) = \sum_{k=0}^{\tau} \beta^k \ell(\mathbf{X}(\tau - k)), \quad (12)$$

with $\beta \in (0, 1)$. The QoS evolves as a linear system,

$$\mathcal{L}(\tau) = \beta \mathcal{L}(\tau - 1) + \ell(\mathbf{X}(\tau)), \quad (13)$$

and through local control is subject to constraints to ensure $\mathcal{L}(\tau) \sim 0$ for each τ .

Here we focus on power deviation: $\ell(\mathbf{X}(k)) = [P(\mathbf{X}(k)) - \mathcal{P}_0]T$, where $P(\mathbf{X}(k))$ is power consumption at state $\mathbf{X}(k)$, \mathcal{P}_0 is the desired long-term average, and T is the sampling interval. Cycling is also considered in prior work.

We require a model for a collection of N loads, rather than an individual, and a model in continuous time. Letting u denote the power deviation from the baseline, and abusing notation slightly to denote $\mathcal{L}(\tau)$ as the sum of QoS over all loads, we approximate to obtain a continuous time model similar to (5):

$$\begin{aligned} \frac{d}{dt}\mathcal{L}(t) &= -\alpha \mathcal{L}(t) - u(t), \quad |\mathcal{L}(t)| \leq C \\ -\eta_- &\leq u(t) \leq \eta_+ \end{aligned}$$

The following conventions are adopted for the charge power limit η_- , discharge power limit η_+ , capacity C , and dissipation rate α :

$$\begin{aligned} \eta_- &= N(\mathcal{P}_m - \mathcal{P}_0), \quad \eta_+ = N\mathcal{P}_0, \\ C &= \varepsilon N \frac{\mathcal{P}_0 T}{1 - \beta}, \quad \alpha = \frac{1}{T} \ln(1/\beta), \end{aligned}$$

where $\varepsilon \in (0, 1)$ is given. The limits η_{\pm} are defined in Section II-D, while α is a result of the continuous-time approximation of (13).

C. Feedforward command signal

The feedforward command signal is obtained from the solution to a quadratic program whose objective is to minimize the cost of supplying forecast net load l with a combination of traditional generation g , and power deviation from flexible loads u . Flexible loads consist of M_t classes of TCLs and M_p classes of non-TCLs.

Quadratic program over time-period $[t, T]$:

$$\begin{aligned}
& \underset{g, x, \mathcal{L}_i(t)}{\text{minimize}} && \int_t^T \left\{ c_0(g(t) - \bar{l}) + \varrho_0\left(\frac{d}{dt}g(t)\right) \right. \\
& && + \sum_{i=1}^{M_t} \left(c_i(x_i(t)) + \varrho_i\left(\frac{d}{dt}x_i(t)\right) \right) \\
& && + \left. \sum_{i=M_t+1}^{M_t+M_p} \left(c_i(\mathcal{L}_i(t)) + \varrho_i\left(\frac{d}{dt}\mathcal{L}_i(t)\right) \right) \right\} dt \\
& \text{subject to} && l(t) = g(t) + \sum_i u_i(t), \\
& && \frac{d}{dt}x_i(t) = -\alpha_i x_i(t) - u_i(t)
\end{aligned} \tag{14}$$

where \bar{l} is the average net load, and the individual cost functions are quadratic:

$$c_i(x) = Q_i x^2, \quad \varrho_i(x) = R_i x^2, \quad i \in \{0, 1, \dots, M_t + M_p\}.$$

Costs associated with generation are assumed to dominate those corresponding to VES resources: $Q_0 \gg Q_i$, $R_0 \gg R_i$. Additional constraints are imposed to ensure that total energy deviation from VES resources is zero:

$$0 = \int_t^T u_i(\tau) d\tau \tag{15}$$

The various constraints involving SoC explained in sections III-A and III-B are regarded as *necessary* conditions for feasibility, since they only impose a constraint on the aggregate. The results in Section II and prior research indicate that these necessary conditions are very nearly sufficient; to ensure sufficiency, it may be necessary to reduce the SoC limit C by a small amount.

Each load class can be assumed homogeneous by design. Even though residential water heaters vary in capacity and time constants, local control loops can be designed so that the aggregate appears as a homogeneous resource [5], [6], [16]. For loads of significantly different behavior, it is necessary to bin into different load classes. For example, in the experiments described in the next section, water heaters were binned into two different classes (residential and commercial).

IV. NUMERICAL EXPERIMENTS

Five classes of electric loads were simulated using the feedforward architecture described in Section III: residential air conditioners (ACs), residential water heaters with faster cycle times (fWHs), commercial water heaters with slower cycle times (sWHs), refrigerators (RFGs), and pool pumps (PPs).

The net load used in our experiments resembles the problematic ‘duck curve’ seen in areas with high solar penetration, as illustrated in Fig. 1. Following these ramps with traditional generation is costly.

The quadratic program (14) was implemented in discrete time with a sampling interval of ten minutes, a time horizon of 24 hours, and solved using MATLAB’s ‘quadprog’ function.

TABLE I
LOAD PARAMETERS - CURRENT SATURATION

Par.	Unit	ACs	fWHs	sWHs	RFGs	PPs
N	million	7	0.7	0.7	17	1.2
η_-	GW	22.23	3.31	3.35	1.96	0.60
η_+	GW	16.98	0.19	0.15	1.45	0.06
α	hours ⁻¹	0.25	0.04	0.01	0.10	0.004
C	GWh	5.6	0.26	0.70	0.43	14.55

A. Balancing California Today

For the results surveyed here, the number of loads per class was chosen based on a residential energy consumption survey of California [17]. Physical parameters from [18] were used for the TCL models. This information is summarized in Table I.

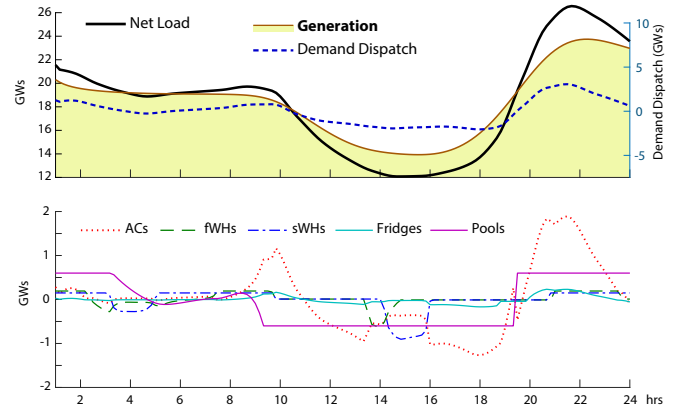


Fig. 8. Optimal power trajectories considering current load profile.

Fig. 8 shows optimal power trajectories over a 24-hour time horizon. Traditional generation and VES combine to optimally match the net load over this time period. The power trajectories represent power supplied to the grid, e.g., when demand dispatch is positive, loads are consuming less than nominal power.

It is evident that load classes cooperate in order to match the net load. Throughout the late morning and early afternoon, the VES systems charge up. In practice, this means houses become slightly cooler than average, and water heaters become slightly hotter than average. Then, during the evening ramp, the VES systems discharge. In practice, this means houses become slightly warmer than average and water heaters become slightly cooler than average. In any case, the use of local control can ensure that temperatures always remain within predefined limits [5], [6], [16]. The value to the grid comes from the combined responses of millions of loads.

As a result of this coordinated effort by VES resources, traditional generation follows a much flatter trajectory.

B. A power hungry future

Currently, about 90% of residential water heaters in California are powered by natural gas [19], but plans are underway in California to transition to electric loads because

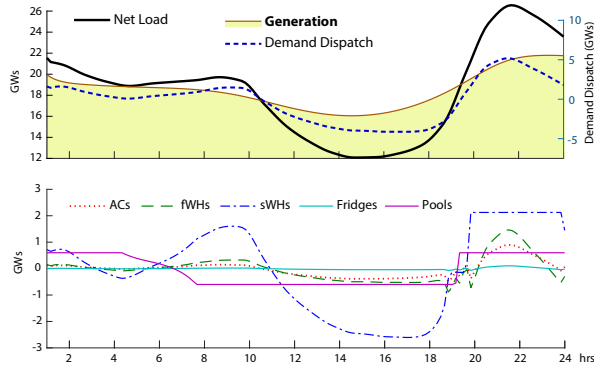


Fig. 9. Optimal power trajectories for California 2020.

TABLE II

LOAD PARAMETERS - OPTIMISTIC SATURATION

Par.	Unit	ACs	fWHs	sWHs	RFGs	PPs
N	million	7	10	10	17	1.2
η_-	GW	22.23	47.25	47.88	1.96	0.60
η_+	GW	16.98	2.75	2.13	1.45	0.06
α	hours ⁻¹	0.25	0.04	0.01	0.10	0.004
C	GWh	5.6	3.75	10.0	0.43	14.55

of abundant electricity, and the evident flexibility of electric water heaters. The results in this section simulate that optimistic future. Parameters are summarized in Table II.

Results from this experiment are shown in Fig. 9. Similar to the previous experiment, VES resources are coordinating in order to match the net load curve. As a result, very little ramping is required from traditional generation. From these results, it is easy to visualize a future where the base net load is nearly constant, and any deviations to the base load can be satisfied by flexible loads.

C. The problem with price signals

The use of price signals is commonly advocated for incentivizing load shifting. This is in fact the definition of “demand response”, according to the FERC website.

In the context of this paper, it is obvious that real-time prices (not posted in advance) are not the right signal. The challenges considered in this paper are largely predictable – large ramps in the morning and afternoon, accompanied by a large peak in power production. There is no economic rationale for withholding information about the load forecast and the concomitant forecast of electricity prices from the consumers.

A popular alternative to real-time prices is the utilization of *time-of-use* (TOU) rates; in this scheme, the prices are published well in advance.

Consider an aggregator that has contracts with a large number of customers. Subject to bounds on QoS for each customer, the aggregator will minimize the overall energy bill for the consumers. This optimization problem is formulated as a quadratic program similar to (14).

Fig. 10 shows the optimal solution obtained in one experiment in which the price is constant, except for a 90 minute period during which it increases by 10% (period marked in

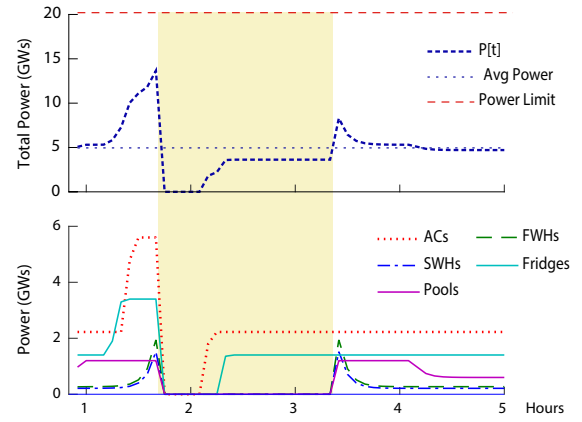


Fig. 10. Optimal response to a price increase over a 90 minute period.

band of yellow). The total power consumption is shown in the upper figure. The well-known “re-bound effect” is clearly seen after the price surge. More pronounced is a “pre-bound”: a large spike in demand just before the event. The BA will not be happy with this outcome.

The nature of the optimal solution is nearly independent of price. An increase of price by 200% will result in a similar set of power trajectories.

The fundamental problem with price signals is that consumers are not interested in power as a commodity. A realistic economic analysis would recognize that consumers are interested in quality of service as defined in this paper, and do not care about how that service is obtained.

Consider Fig. 6, showing the power trajectory for a typical hot water heater. Nobody taking a shower is concerned with the details of this “power spike train”. This is a blessing from the point of view of control since it means consumers are oblivious to substantial changes in power consumption patterns. It is a barrier to those who wish to apply competitive equilibrium theory and auction theory to control the grid through dynamic prices.

V. CONCLUSIONS

The results of this paper provide further evidence that flexible loads can provide reliable grid services that are likely far cheaper than current battery systems. The focus on feedforward control based on load and weather forecasts is motivated by the fact that the loads of interest here are much like batteries, and may need to be “pre-charged” (or discharged) in advance.

The relatively slow time-scale MPC control architecture requires information at the BA from each class of loads in terms of aggregate power consumption and aggregate SoC. It is likely that the ideas of [5] can be extended to eliminate the need for aggregate power consumption measurements – this is a topic of current research.

A question remains regarding how much capacity from the VES resources should be utilized for the feedforward control architecture, so that sufficient capacity is available for real-time feedback control. This and large scale simulations of the entire control architecture are the focus of future research.

REFERENCES

- [1] ASSOCIATED PRESS, Los Angeles Times, “Tesla builds world’s largest battery in Australian outback,” <http://www.latimes.com/business/technology/la-fi-tn-tesla-australia-20171201-story.html>, Dec. 1 2017.
- [2] Lazard, “Lazard’s Levelized Cost of Storage - Version 2.0,” Lazard, Tech. Rep., 2016.
- [3] Y. Chen, A. Bušić, and S. Meyn, “Individual risk in mean field control with application to automated demand response,” in *53rd IEEE Conference on Decision and Control*, Dec 2014, pp. 6425–6432.
- [4] S. Meyn, P. Barooah, A. Bušić, Y. Chen, and J. Ehren, “Ancillary service to the grid using intelligent deferrable loads,” *IEEE Trans. Automat. Control*, vol. 60, no. 11, pp. 2847–2862, Nov 2015.
- [5] J. Mathias, A. Bušić, and S. Meyn, “Demand dispatch with heterogeneous intelligent loads,” in *Proc. 50th Annual Hawaii International Conference on System Sciences (HICSS)*, and *arXiv 1610.00813*, 2017.
- [6] Y. Chen, U. Hashmi, J. Mathias, A. Bušić, and S. Meyn, “Distributed control design for balancing the grid using flexible loads,” in *IMA volume on the control of energy markets and grids*. Springer, 2017.
- [7] H. Hao, B. Sanandaji, K. Poolla, and T. Vincent, “A generalized battery model of a collection of thermostatically controlled loads for providing ancillary service,” in *51st Annual Allerton Conference on Communication, Control, and Computing*, Oct 2013, pp. 551–558.
- [8] H. Hao, B. M. Sanandaji, K. Poolla, and T. L. Vincent, “Aggregate flexibility of thermostatically controlled loads,” *IEEE Trans. on Power Systems*, vol. 30, no. 1, pp. 189–198, Jan 2015.
- [9] P. Kundur, *Power system stability and control*, ser. EPRI power system engineering. McGraw-Hill New York, 1994, vol. 7.
- [10] P. Barooah, A. Bušić, and S. Meyn, “Spectral decomposition of demand-side flexibility for reliable ancillary services in a smart grid,” in *Proc. 48th Annual Hawaii International Conference on System Sciences (HICSS)*, Kauai, Hawaii, 2015, pp. 2700–2709.
- [11] K. Hreinsson and A. Scaglione, “Stochastic power system optimization with responsive thermostatically controlled loads,” *Submitted for publication*, 2018.
- [12] N. O’Connell, E. Hale, I. Doebber, and J. Jorgenson, “On the inclusion of energy-shifting demand response in production cost models: Methodology and a case study,” National Renewable Energy Laboratory (NREL), Golden, CO, United States, Tech. Rep., July 20 2015. [Online]. Available: <https://www.osti.gov/servlets/purl/1215318>
- [13] N. Cammardella, B. Moye, Y. Chen, and S. Meyn, “An energy storage cost comparison: Li-ion batteries vs distributed load control,” in *Clemson University Power System Conference*, Sep 2018, p. PP.
- [14] M. Sadegh, L. Xie, and C. Singh, “Reserves from controllable swimming pool pumps: Reliability assessment and operational planning,” in *51st Hawaii International Conference on System Sciences (HICSS)*, Jan 2018.
- [15] K. J. Astrom and R. M. Murray, “Feedback systems,” *Princeton Univ*, 2016.
- [16] A. Bušić and S. Meyn, “Distributed randomized control for demand dispatch,” in *IEEE Conference on Decision and Control*, Dec 2016, pp. 6964–6971.
- [17] United States Energy Information Administration, *2009 Residential Energy Consumption Survey*.
- [18] J. Mathieu, M. Dyson, D. Callaway, and A. Rosenfeld, “Using residential electric loads for fast demand response: The potential resource and revenues, the costs, and policy recommendations,” in *ACEEE Summer Study on Energy Efficiency in Buildings*. Citeseer, 2012, pp. 189–203.
- [19] C. Palmgren, N. Stevens, M. Goldberg, R. Barnes, and K. Rothkin, “California Residential Appliance Saturation Study,” KEMA, Inc., Tech. Rep., 2009, <http://www.energy.ca.gov/2010publications/CEC-200-2010-004/CEC-200-2010-004-ES.PDF>.